Probability estimates of heavy precipitation events in a flood-prone central-European region with enhanced influence of Mediterranean cyclones

J. Kysely¹ and J. Picek²

¹Institute of Atmospheric Physics AS CR, Prague, Czech Republic
²Technical University, Liberec, Czech Republic

Abstract. Due to synoptic-climatological reasons as well as a specific configuration of mountain ranges, the northeast part of the Czech Republic is an area with an enhanced influence of low-pressure systems of the Mediterranean origin. They are associated with an upper-level advection of warm and moist air and often lead to heavy precipitation events. Particularities of this area are evaluated using a regional frequency analysis. The northeast region is identified as a homogeneous one according to tests on statistical characteristics of precipitation extremes (annual maxima of 1- to 7-day amounts), and observed distributions follow a different model compared to the surrounding area. Noteworthy is the heavy tail of distributions of multi-day events, reflected also in inapplicability of the L-moment estimators for the general 4-parameter kappa distribution utilized in Monte Carlo simulations in regional homogeneity and goodness-of-fit tests. We overcome this issue by using the maximum likelihood estimation. The Generalized Logistic distribution is identified as the most suitable one for modelling annual maxima; advantages of the regional over local approach to the frequency analysis consist mainly in reduced uncertainty of the growth curves and design value estimates. The regional growth curves are used to derive probabilities of recurrence of recent heavy precipitation events associated with major floods in the Odra river basin.

1 Introduction

The paper deals with the frequency modelling of heavy precipitation events in a flood-prone region in the northeast part of the Czech Republic, close to the borders with Poland and Slovakia (Fig. 1). A distinct feature of this area consists in a frequent occurrence, relative to surrounding parts of central Europe, of heavy multi-day rainfall events leading to floods (Šamaj et al., 1983; Štekl et al., 2001). This is related to an enhanced influence of low-pressure systems of the Mediterranean origin (mostly travelling from the Adriatic region northeastward; weather situation “Zugstrasse Vb” – cf. Mudelsee et al., 2004; Kundzewicz et al., 2005), which are associated with an upper-level advection of warm and moist air. Two main mechanisms support the occurrence of heavy rainfall (or snowfall) in this region: First, trajectories of the cyclones, moving towards north or northeast, often pass in the vicinity of the area, and sometimes the cyclones become quasi-stationary or retrograde. Their centres frequently persist (for several days) northeast of the region over south Poland, and the associated fronts and cloud belts produce multi-day rainfall episodes. Second, a specific configuration of mountain ranges in the area under study (reaching almost 1500 m a.s.l. in the two main mountain ranges, Jeseníky Mts. and Beskydy Mts., and overtopping the surrounding lowland areas by more than 1000 m) supports uplifts and convergence of the moist air (due to “funnel effects”; e.g. Štekl et al., 2001).

The distinct regional features linked to the pronounced influence of the Mediterranean cyclones are reflected also in statistical characteristics of the frequency distributions of rainfall extremes. A recent study (Kyselý and Picek, 2007) showed that the peculiarities of this region consist in

- a different frequency model suitable for precipitation extremes instead of the commonly utilized Generalized Extreme Value (GEV) distribution,
- very heavy tails of distributions of annual maxima of multi-day precipitation totals (according to tests on the tail index), which result also in inapplicability of L-moment estimators for the 4-parameter kappa distribution (Hosking, 1994) in regional homogeneity tests and goodness-of-fit tests.

Published by Copernicus Publications on behalf of the European Geosciences Union.
The aims of this study are to develop frequency models and to improve estimates of probabilities of heavy rainfall in this flood-prone area under a regional approach (Hosking and Wallis, 1997; Sect. 2), using a high-quality dataset with a much better spatial coverage (Sect. 3). The issue of inapplicability of the L-moments for the kappa distribution is overcome by the use of the maximum likelihood estimation (Sect. 4.1); regional homogeneity of the new dataset is verified (Sect. 4.2); the most appropriate probability distribution of precipitation maxima is selected according to a goodness-of-fit test (Sect. 4.3); and regional growth curves and design values are constructed, together with estimates of their uncertainty based on a parametric bootstrap (Sect. 4.4).

2 Methods

Regional frequency analysis, widely used in hydrology (e.g., Pilon and Adamowski, 1992; Adamowski, 2000; Kjeldsen and Rosbjerg, 2002) as well as climatology (e.g. Guttmann et al., 1993; Smithers and Schulze, 2001; Gellens, 2002; Fowler and Kilsby, 2003; Boni et al., 2006), is utilized to examine high quantiles (design values) of heavy 1-day and multi-day precipitation amounts. The method consists in delineating fixed regions that are homogeneous according to statistical characteristics of the probability distributions of extremes, i.e. the at-site distributions are identical except for a site-specific scaling factor. The examined area forms one of the 4 homogeneous regions in the Czech Republic (Kysely et al., 2007; the study made use of a subset of the data over a shorter time period, so the regional homogeneity is revisited in Sect. 4.2). The general advantage of the regional approach compared to a single-site analysis is that more data are available for the probability modelling, including the choice of the frequency distribution and estimation of its parameters and design values, and the estimates become more reliable and spatially more coherent.

The Lu-Stedinger test (Lu and Stedinger, 1992) and Hosking-Wallis test (Hosking and Wallis, 1997) are applied to evaluate whether the basic assumption of the procedure is valid and the present dataset forms a homogeneous region. The tests compare at-site estimates with a regional estimate of a quantity that measures some aspect of the frequency distribution of maxima, and involve Monte Carlo simulations of homogeneous regions with the general 4-parameter kappa distribution (Hosking, 1994) in order to obtain the mean and standard deviation of a chosen dispersion measure (for details see Kysely and Picek, 2007).

The goodness-of-fit of several candidate probability distributions is then assessed in terms of the difference between the L-kurtosis of the fitted distribution and the regional average L-kurtosis (Hosking and Wallis, 1997). The significance of this difference depends on the sampling variability of the regional average L-kurtosis, which is again obtained by simulations of homogeneous regions with the kappa distribution. L-moments are used for the estimation of all distributions, except for the cases of the kappa distribution discussed below when they are replaced by the maximum likelihood estimators (Park and Jung, 2002).

More details on the statistical methods for the regional frequency analysis may be found in Hosking and Wallis (1997).
Table 1. Values of the test statistics of the regional homogeneity tests. \( R_k \) \( (k=1, 2, 3, 5, 7) \) stands for annual maxima of \( k \)-day precipitation amounts. The test result indicating heterogeneity is shown in bold.

<table>
<thead>
<tr>
<th>variable</th>
<th>Lu-Stedinger test</th>
<th>Hosking-Wallis test based on L-moments</th>
<th>Hosking-Wallis test based on MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>56.60</td>
<td>0.81</td>
<td>−0.30</td>
</tr>
<tr>
<td>R2</td>
<td>43.51</td>
<td>−0.62</td>
<td>−1.35</td>
</tr>
<tr>
<td>R3</td>
<td>35.19</td>
<td>( x^a )</td>
<td>−2.07</td>
</tr>
<tr>
<td>R5</td>
<td>32.47</td>
<td>( x^a )</td>
<td>−2.09</td>
</tr>
<tr>
<td>R7</td>
<td>25.98</td>
<td>( x^a )</td>
<td>−1.95</td>
</tr>
</tbody>
</table>

\( a \) The Hosking-Wallis test with L-moment estimators of the kappa distribution was impracticable due to the parameter space restriction.

Table 2. Values of the Z-statistics of the goodness-of-fit test for candidate 3-parameter distributions. Distributions rejected at the 0.10 (0.05) level are labelled with * (**).

<table>
<thead>
<tr>
<th>variable</th>
<th>Generalized Extreme Value (GEV)</th>
<th>Generalized Logistic (GLO)</th>
<th>Lognormal</th>
<th>Pearson type III</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>−0.85</td>
<td>1.32</td>
<td>−2.02**</td>
<td>−4.08**</td>
</tr>
<tr>
<td>R2</td>
<td>−1.40</td>
<td>0.10</td>
<td>−2.79**</td>
<td>−5.18**</td>
</tr>
<tr>
<td>R3( a )</td>
<td>−1.30</td>
<td>−0.44</td>
<td>−2.59**</td>
<td>−4.78**</td>
</tr>
<tr>
<td>R5( a )</td>
<td>−2.31**</td>
<td>−1.46</td>
<td>−3.63**</td>
<td>−5.89**</td>
</tr>
<tr>
<td>R7( a )</td>
<td>−0.92</td>
<td>0.05</td>
<td>−2.04**</td>
<td>−3.94**</td>
</tr>
</tbody>
</table>

\( a \) Maximum likelihood estimators of the kappa distribution were used because sample L-moment ratios violate restricted parameter space conditions (see Sect. 4.1).

3 Data

Annual maxima of 1- to 7-day precipitation amounts are available at 33 rain-gauge stations in the northeast region of the Czech Republic (Fig. 1). Only stations with complete records over at least 31 consecutive years, and without significant station moves are included in the dataset. The maximum length of the samples is 45 years (period 1961–2005). The stations approximately evenly cover the region, and their altitudes range from 220 to 1490 m a.s.l. The dataset is superior to the one employed in Kysely and Picek (2007) as it involves a much larger number of sites in the region (33 compared to 12) and extends to the very recent past (2005).

Note that the multi-day amounts are particularly important since larger-scale impacts from heavy precipitation are mostly due to multi-day episodes, and 5-day amounts are frequently considered measures of the occurrence of floods in Regional Climate Model studies (e.g. Christensen and Christensen, 2003; Gao et al., 2006).

4 Results

4.1 Maximum likelihood estimators for the kappa distribution

For the estimation of the four parameters of the kappa distribution in simulations of homogeneous regions in the regional homogeneity and goodness-of-fit tests, the method of L-moments (Hosking, 1990) is conventionally used. Even though it is robust, the parameter space is restricted to ensure the existence of the L-moments and the uniqueness of parameters. Therefore the computation routine sometimes fails to get valid estimates for some datasets, because no unique solution exists inside the restricted region (cf. Park and Jung, 2002; Kysely and Picek, 2007). In the present application, the original Hosking-Wallis tests (as well as goodness-of-fit tests) with the kappa distribution were impracticable for 3- to 7-day precipitation amounts since sample L-moment ratios estimated from the data do not correspond to the parameter space that ensures the existence of the L-moments and the uniqueness of parameters.

The issue was surmounted by the application of the maximum likelihood estimation with the kappa distribution. The maximum likelihood estimates are obtained by minimizing negative log-likelihood function. Since no explicit minimizer is possible, the function can be numerically minimized by
4.2 Tests on regional homogeneity

Values of the regional homogeneity test statistics are given in Table 1. The Lu-Stedinger test rejects the null-hypothesis that the region is homogeneous at the 0.05 level if the test statistic exceeds the 95th percentile of the χ² distribution with N−1 degrees of freedom (equal to 46.19 in the present application; N denotes the number of stations forming the region). The Hosking-Wallis test yields “possible heterogeneity” (“definite heterogeneity”) if the value of the test statistic is ≥1 (≥2). The test results show that the region is homogeneous in all examined variables except for 1-day annual maxima for which a slight heterogeneity is indicated by the Lu-Stedinger test. However, since the region is homogeneous for 1-day events according to the Hosking-Wallis test, we preserve the same grouping for all examined variables (1- to 7-day annual maxima). The Hosking-Wallis tests were impracticable with L-moment estimators of the kappa distribution for 3- to 7-day amounts; if maximum likelihood estimators are employed, the regional homogeneity for all datasets is supported.

4.3 Goodness-of-fit tests

Results of the goodness-of-fit tests for candidate 3-parameter distributions are summarized in Table 2; a distribution is rejected at the 0.10 (0.05) level if |Z| > 1.64 (|Z| > 1.96). For all examined variables (1- to 7-day precipitation amounts), the lognormal and Pearson type III distributions are rejected (the latter was used in previous studies on probabilities of heavy rainfall in the area under study, following a general recommendation of WMO). The Generalized Extreme Value (GEV) and Generalized Logistic (GLO) distributions (see Appendix A for the description of the latter) are two alternative models supported by the goodness-of-fit test; however, the GEV is rejected at the 0.05 level for 5-day amounts, and the absolute values of the test statistics tend to be larger for GEV than GLO. This is why the GLO appears as the most suitable frequency model for annual precipitation maxima in the region. The finding is in accord with Kyselý and Picek (2007) but it has received a better support: data from 33 stations over 1961–2005 enter the goodness-of-fit tests (compared to 12 stations over 1961–2000; the number of station-years has increased from 480 to 1443), and the tests for multi-day events with the maximum likelihood estimation for the kappa distribution clearly show superiority of the GLO distribution (originally they remained impracticable with the L-moments).

It should be noted that the GLO (termed also log-logistic) distribution is a model which originates in hydrology (Ahmad et al., 1988), and although it is a useful alternative for maxima of daily precipitation amounts in some regions (see also Fitzgerald, 2005), it has been employed rarely in practical applications. We utilize a reparametrized version of the log-logistic distribution of Ahmad et al. (1988) in which the parameters are analogous to those of the GEV distribution. The GEV and GLO distributions are closely related 3-parameter models that rank among distributions with the same weight of the upper tails, but the latter enables somewhat better representation of long heavy tails and outlying very extreme observations. A comparison of their probability density functions for 5-day precipitation maxima is shown in Fig. 2; dimensionless parameters were estimated using the L-moment regional procedure (Sect. 4.4).

We also remark that the GLO distribution is suitable specifically for the area under study where distributions of precipitation maxima possess heavy upper tails due to the enhanced influence of Mediterranean cyclones. Unlike the GEV distribution, which appears to be a rather universal model for rainfall extremes in other parts of central Europe (e.g. Kyselý and Picek, 2007; Gaál, 2006), the GLO distribution is not applicable in any other region of the Czech Republic according to the goodness-of-fit tests.
Fig. 3. At-site (left) and regional (right) estimates of 50-yr return levels of 1-day precipitation amounts from the Generalized Logistic distribution. The vertical bars show 90% confidence intervals estimated using a parametric bootstrap, taking intersite dependence into account in the regional approach. The stations are ranked with respect to their altitudes.

Fig. 4. Same as in Fig. 3 except for 5-day precipitation amounts.

4.4 Regional growth curves and estimates of high quantiles

Parameters of the regional distributions were estimated by means of a regional algorithm based on L-moments (Hosking and Wallis, 1997). Sample L-moment ratios at individual sites were combined to give regional average L-moment ratios; analogously to the method of conventional moments, the first three L-moments were used to obtain parameters of a given distribution. Mean of the at-site samples was taken as the scaling factor (index storm). The accuracy of estimates was determined from a parametric bootstrap, taking into account the intersite dependence in terms of correlation matrices.

Comparison of at-site and regional estimates of 50-yr return values of 1-day and 5-day precipitation amounts based on the GLO distribution (Figs. 3 and 4) shows main bene-
Table 3. Return periods associated with the July 1997 record-breaking 5-day totals at selected sites, estimated using the regional approach and the GLO distribution.

<table>
<thead>
<tr>
<th>station</th>
<th>Fulnek (290 m a.s.l.)</th>
<th>Valašské Meziříčí (334 m a.s.l.)</th>
<th>Rajnochovice (405 m a.s.l.)</th>
<th>Karlovice (505 m a.s.l.)</th>
<th>Praděd&lt;sup&gt;a&lt;/sup&gt; (1490 m a.s.l.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>observed maxima</td>
<td>292.1</td>
<td>376.3</td>
<td>433.0</td>
<td>320.4</td>
<td>455.0</td>
</tr>
<tr>
<td>July 1997 [mm]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>return period</td>
<td>330</td>
<td>410</td>
<td>490</td>
<td>280</td>
<td>250</td>
</tr>
<tr>
<td>[years]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd largest</td>
<td>112.6</td>
<td>142.8</td>
<td>161.3</td>
<td>127.6</td>
<td>185.2</td>
</tr>
<tr>
<td>observation since</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1961 [mm]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> The station terminated operation on 31 August 1997.

Estimates of return periods of precipitation extremes observed in July 1997 (the region covers the area most affected with heavy precipitation leading to the worst flood in the Odra river basin at least since the late 19th century, and the most disastrous in the Czech Republic during the 20th century) are based on the GLO distribution; the GEV distribution was employed for comparison of the two statistical models. A characteristic feature of this event was its multi-day nature; maximum 5-day amounts recorded at individual sites were between 173 and 455 mm. The regional analysis yields estimates of their return periods between 55 and 490 years with the GLO distribution (the range is over measuring sites, with the mean of 167 years). The GEV distribution seems to overestimate the upper range over the sites as it leads to return periods of 54 to 640 years (with the mean of 191 years). The relatively large variation of estimates of return periods under the regional approach is expected since observations at all sites in the region are involved in the estimation of the probability distribution at a particular site (the at-site analysis would yield distributions that reflect much smaller samples, and a larger weight would be given to the observed maxima at a site). Note that the very long return periods estimated for the 1997 event at several sites seem entirely reasonable after data are scrutinized; the second largest observations over the period of record (1961–2005) were much lower as shown in Table 3. The range of return periods associated with the July 1997 heavy 5-day precipitation is between 80 and 250 years at two thirds of stations.

5 Conclusions

The study presents main findings concerning the frequency modelling and design value estimation of heavy 1-day and multi-day precipitation events in a specific flood-prone region in central Europe, in which an enhanced influence of cyclones of the Mediterranean origin plays a crucial role in producing precipitation extremes. We modify the standard L-moment based algorithm of the regional frequency analysis in that the maximum likelihood is used to estimate the 4-parameter kappa distribution in the regional homogeneity tests and the goodness-of-fit tests when necessary. This is because L-moments do not provide unambiguous solutions of the kappa distribution parameters in particular cases, e.g. when the upper tails are very heavy. The Generalized Logistic (GLO) distribution is found to be an appropriate model for the annual precipitation maxima in the region; it gives a better fit to the data than the more commonly applied Generalized Extreme Value (GEV) distribution. Design values of the 1-day and multi-day rainfall amounts, corresponding to return periods of 10- to 200-ys, are derived together with their uncertainties. It is shown that the regional approach leads to superior results compared to an at-site (local) analysis, and the return periods of the heavy 5-day amounts that caused massive floods on the Odra river in July 1997 are between 80 and 250 years at most sites.

A similar framework may be useful in studying probabilities of heavy rainfall in other parts of the world, and the GLO distribution may be a model suitable for heavy-tailed data in conditions when no other 3-parameter frequency distribution
Thanks are due to P. Štepánek (Czech Hydrometeorological Institute, Brno) for providing the rainfall data and performing quality checks, and L. Gaál (Slovak Technical University, Bratislava) and J. Hošek (Institute of Atmospheric Physics, Prague) for assistance in drawing figures. The study was supported by the Grant Agency of AS CR under project B300420601.

Appendix A

Generalized Logistic (GLO) distribution

The cumulative distribution function of the GLO distribution with parameters $\xi$ (location), $\alpha$ (scale) and $k$ (shape) is (Hosking and Wallis, 1997)

$$F(x) = \frac{1}{1 + e^{-\frac{x-\xi}{\alpha}}},$$

where

$$y = -\frac{\ln \left(1 - \frac{k(x-\xi)}{\alpha}\right)}{k}, \quad k \neq 0,$$

$$y = \frac{x - \xi}{\alpha}, \quad k = 0.$$

The distribution is bounded at $\xi + \frac{2}{k}$ from right (left) if $k > 0$ ($k < 0$). L-moments are defined for $-1 < k < 1$; the first three population L-moments are

$$\lambda_1 = \xi + \alpha \left(1 - \frac{\pi}{\sin k\pi}\right),$$

$$\lambda_2 = \frac{\alpha k \pi}{\sin k\pi},$$

and

$$\lambda_3 = -k\lambda_2.$$

The method of L-moments fits the GLO distribution by choosing its parameters so that the first three L-moments $\lambda_1$, $\lambda_2$, $\lambda_3$ match the corresponding sample estimates $l_1$, $l_2$, $l_3$. The L-moment estimators of $k$, $\alpha$ and $\xi$ are given by

$$k = -\frac{l_3}{l_2},$$

$$\alpha = \frac{l_2 \sin k\pi}{k\pi},$$

and

$$\xi = l_1 - \alpha \left(1 - \frac{\pi}{k\sin k\pi}\right).$$

Acknowledgements. Thanks are due to P. Štepánek (Czech Hydrometeorological Institute, Brno) for providing the rainfall data and performing quality checks, and L. Gaál (Slovak Technical University, Bratislava) and J. Hošek (Institute of Atmospheric Physics, Prague) for assistance in drawing figures. The study was supported by the Grant Agency of AS CR under project B300420601.

Edited by: P. Alpert, H. Saaroni, and E. Heiftz  
Reviewed by: two anonymous referees

References


